Problem1:

1. **From normalized formulas**

Mean: 1.0489703904839582

Variance: 5.4272206818817255

Skewness: 0.8792880598472457

Kurtosis: 23.06998251061053

2. **From statistical package**

Mean: 1.0489703904839585

Variance: 5.4217934611998455

Skewness: 0.8806086425277364

Kurtosis: 23.122200789989723

3. **Either biased or unbiased**

H0: The statistical package is unbiased

H1: The statistical package is biased

Then perform the t-tests for mean and variance and z-test for skewness and kurtosis under different number of sample sizes, and compare the resulted p-value with 0.05, if $p > 0.05$, then fail to reject H0; else reject H0. And as shown below, the mean and variance values are biased calculated, skewness is unbiased, and as the sample size growing, the kurtosis calculation moved from unbiased to biased.

Results for Sample Size: 100

Reject H0 for Mean.

Reject H0 for variance.

Fail to reject H0 for skewness.

Fail to reject H0 for kurtosis.

Results for Sample Size: 1000

Reject H0 for Mean.

Reject H0 for variance.

Fail to reject H0 for skewness.

Fail to reject H0 for kurtosis.

Results for Sample Size: 100000

Reject H0 for Mean.

Reject H0 for variance.

Fail to reject H0 for skewness.

Reject H0 for kurtosis.

Results for Sample Size: 1000000

Reject H0 for Mean.

Reject H0 for variance.

Fail to reject H0 for skewness.

Reject H0 for kurtosis.

	Sample Size	P-value Mean	P-value Variance	P-value Skewness \
0	100	NaN	NaN	0.995698
1	1000	NaN	NaN	0.986398
2	100000	NaN	NaN	0.864628
3	1000000	NaN	NaN	0.589800

	P-value Kurtosis
0	0.915114
1	0.736066
2	0.000750
3	0.000000

Problem2:

1.

OLS Results:

Beta values (OLS):

[-0.08738446 0.7752741]

Sigma: 1.008813058320225

MLE Model:

Beta values (MLE): [-0.08738448 0.77527408]

Sigma: 1.0037563087311683

As we can see from the results above, the betas numbers are almost the same, but the sigma value varies, the Sigma (ols) = 1.0088 while the Sigma (mle) = 1.00375, which is because the MLE model is not unbiased.

2.

T-distribution MLE Results:

Intercept: -0.0878, Beta: 0.7737, df: 384.4406, Sigma: 1.1906

AIC for T-distribution model: 545.8889

BIC for T-distribution model: 559.0821

Normal distribution MLE Results:

Intercept: -0.0874, Beta: 0.7753, Sigma: 1.0038

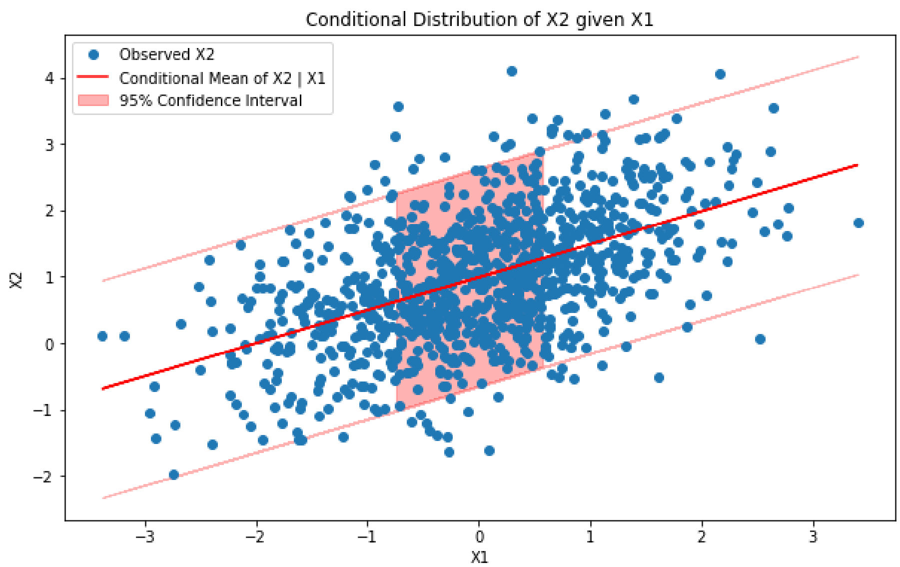
AIC for normal distribution model: 575.0751

BIC for normal distribution model: 584.9701

The T-distribution model provides a better fit based on AIC.

The T-distribution model provides a better fit based on BIC.

While the AIC of T-distribution (545.8889) < AIC of normal distribution MLE (575.0751), so T-distribution fits better; same on the BIC's perspective, the BIC of T-distribution (559.0821) < BIC of normal distribution MLE (584.9701), so T-distribution fits better. Overall, T-distribution is a better fit.


Conditional Distribution of X2 given X1
Observed X2
Conditional Mean of X2 | X1
95% Confidence Interval
X2
X1
4
3
2
1
0
−1
−2
−3
−2
−1
0
1
2
3

3.

Problem 2.4:

$$Y_{n\times 1} = X_{n\times p}\beta_{p\times 1} + \varepsilon_{n\times 1}, \quad \varepsilon \sim N(0, \delta^2)$$

$$\varepsilon_i = (y_i - x_i\beta_i) \sim N(0, \delta^2).$$

$$f(x) = \frac{1}{\sqrt{2\pi\delta^2}} \exp\left(-\frac{x^2}{2\delta^2}\right).$$

$$\Rightarrow f(\varepsilon_i) = \frac{1}{\sqrt{2\pi\delta^2}} \exp\left(-\frac{(y_i - x_i\beta_i)^2}{2\delta^2}\right).$$

Likelihood function is:

$$L(\beta, \delta^2 | Y) = \prod_{i=1}^{n} \frac{1}{\sqrt[n]{2\pi\delta^2}} \exp\left(-\frac{(y_i - x_i\beta_i)^2}{2\delta^2}\right)$$

$$= \frac{1}{\sqrt[n]{2\pi\delta^2}} \exp\left(-\frac{1}{2\delta^2}\sum_{i=1}^{n}(y_i - x_i\beta_i)^2\right).$$

$$= \frac{1}{\sqrt[n]{2\pi\delta^2}} \exp\left(-\frac{1}{2\delta^2}(Y - X\beta)^T(Y - X\beta)\right).$$

$$\Rightarrow \ell(\beta, \delta^2 | Y) = \log\left(\frac{1}{\sqrt[n]{2\pi\delta^2}}\right) + \left(-\frac{1}{2\delta^2}(Y - X\beta)^T(Y - X\beta)\right).$$

$$= -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log(\delta^2) - \frac{1}{2\delta^2}\left((Y - X\beta)^T(Y - X\beta)\right)$$

$$\therefore \frac{\partial \ell}{\partial \beta} = -\frac{1}{\delta^2} X^T(Y - X\beta) = 0$$

$$\therefore X^T(Y - X\hat{\beta}) = 0.$$

$$\therefore X^TY - X^TX\hat{\beta} = 0.$$

$$\therefore X^TX\hat{\beta} = X^TY.$$

$$\therefore \hat{\beta} = (X^TX)^{-1}X^TY.$$

$$\frac{\partial \ell}{\partial \delta^2} = -\frac{n}{2\delta^2} + \frac{1}{2\delta^4}\left((Y - X\beta)^T(Y - X\beta)\right) = 0.$$

$$\therefore \frac{n}{2\delta^2} = \frac{1}{2\delta^4}\left((Y - X\beta)^T(Y - X\beta)\right)$$

$$\therefore \hat{\delta}^2 = \frac{1}{n}(Y - X\hat{\beta})^T(Y - X\hat{\beta})$$

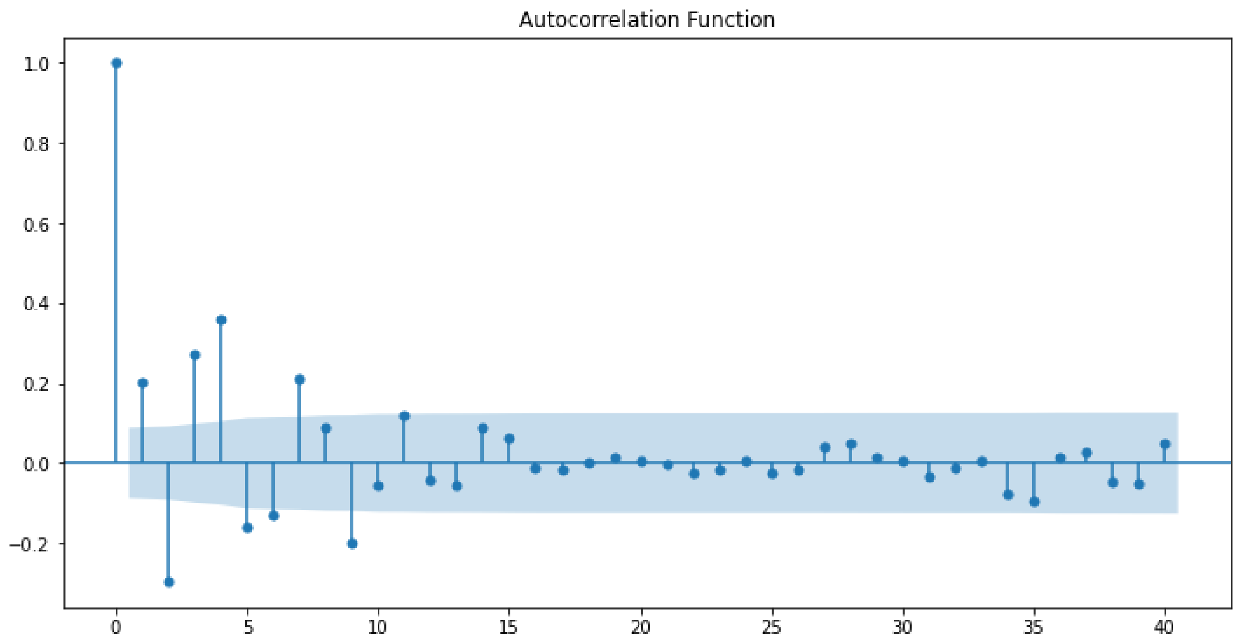
Thus: $\hat{\beta} = (X^TX)^{-1}X^TY.$

$$\hat{\delta}^2 = \frac{1}{n}(Y - X\hat{\beta})^T(Y - X\hat{\beta}).$$

4.

Problem3:

Here we first need to generate an autocorrelation model to check the lags and see which model we need to apply on. As you can see from the chart below, lag1 shows a strong positive autocorrelation (1.0) that means the current value of x is highly correlated with the past values and a strong short-term dependency of the given data. Lag2 decreased significantly but still remains positive, however, lag3 provides a negative correlation. As we continue observing beyond lag3, the autocorrelation becomes smaller and smaller and gradually oscillate around0, which indicates that as we have longer lags, there aren't any correlations exist. So I think the AR model suits better since it can better examine the short-term of the data, so AR(1) or AR(2) might be good fits; however, AR(3) can also fits good cause there're some fluctuations between lag2 and lag3.



Then we write out the code and generate the result below:

	Model	AIC
0	AR(1)	1644.655505
1	AR(2)	1581.079266
2	AR(3)	1436.659807
3	MA(1)	1567.403626
4	MA(2)	1537.941206
5	MA(3)	1536.867709

Best model based on AIC: AR(3)

Since AR(3) has the lowest AIC, then the best model is AR(3).